

Online and Unsupervised Face Recognition for Humanoid Robot: Toward Relationship with People

Lijin Aryananda

Artificial Intelligence Laboratory
Massachusetts Institute of Technology
200 Technology Square Rm 936 Cambridge MA 02139
lijin@ai.mit.edu

1. Introduction

The ability to recognize and remember individuals is crucial and has important implications for the evolution of animal social behavior, particularly complex interactions within groups. Male dolphins have been found to form coalitions, where each group possesses a fertile female. Observation of behavior within the coalitions indicates complex social behavior where dolphins often form ‘coalitions of coalitions’, but each sub-coalition mates only with its own female. This implies the existence of complex social interaction, such as preferential treatment, cooperative behavior, and reciprocity [2]. Such a relationship demands the ability to distinguish conspecific group members as individuals and as kin, remember their relative ranks and past affiliations, and in some cases, remembers the personal histories of help given and received from others [3].

This paper addresses the question of how one may implement such social competence in a humanoid robot. The first part of this task has been tackled by a substantial amount of research in person identification technology using various modalities, including face, body, and speaker recognition [4], [9], [19]. Most of these works attempt to solve the identification problem: given a set of labeled training data and a set of test data, find the correct person label for the test data. We would like to focus and draw attention to how the training data may be acquired in the first place. In most existing face recognition systems, the training images are collected and labeled manually. We argue that this imposes a limitation on the second part of the task: the ability to link contextual information (past behavior, affiliations, etc) with recognition memory of one’s appearance. The contextual knowledge about other people that we acquire through our daily social experience is so rich and complex that manually encoding it into a database along with the corresponding person’s facial images for the robot to memorize is very limiting. We propose that the robot must learn about individuals and their various

characteristics through embodied social interaction, thus directly perceiving the richness of the environment.

As an initial attempt toward this goal, we implemented an online and unsupervised face recognition system, where the robot opportunistically collects, labels, and learns various faces while interacting with people, starting from an empty database. In the rest of the paper, we describe the underlying motivation for this work and discuss related works. We briefly describe Kismet, our robotic platform. We then outline design issues and present an implementation of an online and unsupervised face recognition system for our robot, using the eigenfaces technique [5]. Experiments were performed to examine system behavior. We report details on experiment setting and results. Lastly, we introduce future works to implement other modalities and precursors based on the lessons learned from this implementation.

2. Motivation and Background

The notion of people as distinct individuals plays a very important role in our daily social life. Individual recognition is also a widely reported phenomenon in the animal world, where it contributes to successful maternal interaction, parental care, group breeding, cooperation, and mate choice. If a robot has the ability to recognize and remember people it interacts with, it will be able to learn about characteristics of each individual through interacting with them. This leads to complex social behavior, such as cooperation, dislike, loyalty, and affection. As proposed by [6], if robots have long-term contact with humans, it may be desirable to have them develop individual relationships, which is exactly the aftermath of this social dynamic. Moreover, the ability to distinguish among people allows the robot to build toward more complex social competencies where the idea of people as distinct individuals is crucial, including theory of mind and social referencing.

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 2001		2. REPORT TYPE		3. DATES COVERED 00-00-2001 to 00-00-2001	
4. TITLE AND SUBTITLE Online and Unsupervised Face Recognition for Humanoid Robot: Toward Relationship with People				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Massachusetts Institute of Technology, Computer Science and Artificial Intelligence Laboratory, 32 Vassar Street The Strata Center, Building 32, Cambridge, MA, 02139				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES The original document contains color images.					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 8	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

This work is an extension of an effort to explore socially situated forms of learning for humanoid robots, where the users teach the robot as they would another person. [1] has argued for the many advantages social cues and skills may offer robots that learn from people. Infant-level social competencies have been implemented on our robot, Kismet, as building blocks for exploring socially situated learning between Kismet and its human caregivers. Within this framework, the ability to recognize and remember individuals would expand Kismet's social repertoire, allowing it to develop complex social dynamics, such as attachment to caregivers, dislike of certain people, etc. Teaching the robot would also be easier once the robot is 'familiarized' with the caregivers.

Among the various sensory cues humans use to recognize others, the face is arguably the most important and salient feature. Faces play a very central role in human's social interaction, providing critical information about a person's age, gender, emotional state, etc. Studies have shown that shortly after birth, infants display a clear visual preference for faces [18]. [17] demonstrated that infants could recognize their mother's faces after only 4 hours of discontinuous exposure. Humans exhibit extraordinary analysis and retention of facial images. The ability to learn to recognize faces in an online and unsupervised manner will allow the robot to acquire complex contextual information about individuals and ground this knowledge to some 'meaningful' representations for the robot. For example, the robot may learn to correlate individuals with the robot's emotional state during interaction. Thus, the robot may start to 'like' certain people because somehow when they are around, the robot happens to be happy.

3. Related Work

Research in person identification technology has recently received significant attention, due to the wide range of biometric, information security, law enforcement applications, and Human Computer Interaction (HCI). Face recognition is the most frequently explored modality and has been implemented using various approaches [5]. [9] proposed to combine face and body recognition. Speech recognition has also been widely investigated [19]. The use of multiple modalities has been observed by [7], [8].

[10] presented an online supervised learning method for face recognition, allowing machines to learn directly from sensory input streams. The system receives a video camera output as well as a simulated

auditory sensor. Each individual is labeled by entering the person's name and gender during training session. Both cameras and subjects are static. A recognition accuracy of 95.1% has been achieved on 143 people. The issue of direct coupling between the face recognition system and sensory input is very relevant to our work, due to the requirement of an embodied setting.

[16] proposed an image and video indexing approach. Using a neural network based face detector, extracted faces are grouped into clusters by a combination of a face recognition method using pseudo two-dimensional Hidden Markov Models and a k-means clustering algorithm. The number of clusters is specified manually. Each resulting cluster consists of the face images of one person. Experiments on a TV broadcast news sequence demonstrated that the system is able to discriminate between three different newscasters and an interviewed person. In contrast to this work, we require the recognition system to perform in an automatic and unsupervised manner. Thus, the number of clusters i.e. the number of individuals interacting with the robot at any given time is unknown.

4. Robot Physicality



Figure 1. Kismet is an expressive robotic creature designed for natural social interaction with human[1].

Kismet is an expressive robotic creature with perceptual and motor modalities tailored to natural human communication channels. Kismet has three degrees of freedom to control gaze direction, three degrees of freedom to control its neck, and fifteen degrees of freedom in other expressive components of the face (such as ears, eyebrows, lips, and eyelids). Kismet displays a wide assortment of facial expressions, which mirrors its affective state and produces numerous facial displays for other communicative purposes (see figure 1).

To interact with its caregivers, Kismet uses four color CCD cameras and an unobtrusive wireless microphone. All cameras move with respect to the

head. The positions of the neck and eyes are important both for expressive postures and for directing the cameras towards behaviorally relevant stimuli. [1] found that the manner in which the robot moves eyes and directs its gaze has profound social consequences when engaging people, beyond just steering its cameras to look at interesting things. The hardware and software control architectures have been designed to meet the challenge of real-time processing of visual signals (approaching 30 Hz) and auditory signals (frame size of 10 ms) with minimal latencies (500 ms). Kismet's vision system is implemented on a network of nine 400 MHz commercial PCs running the QNX real-time operating system. Kismet's emotion, behavior, and expressive systems run on a collection of four Motorola 68332 processors. The speech module, including synthesizer and speech processing software runs on Windows NT and Linux.

5. Design Issues and Strategy

5.1 Performance Criteria

Current state of the art in face recognition technology allows for a recognition accuracy of 95% on more than 1000 frontal mugshot-like images when taken in the same day and 80% when taken with a different camera and lighting condition [11]. Our performance criteria, however, is less ambitious in terms of recognition accuracy per image given various constraints described below. For example, in order to accommodate social learning, it would be advantageous for Kismet to be able to distinguish between familiar caregivers and strangers. It would also be useful if Kismet can recognize and thus avoid certain people who often over-stimulate or invade its personal space. Thus, our performance criteria are for the system to be able to recognize and remember people who are relevant to Kismet and interact with it on a regular basis. Keep in mind that we also do not remember every single person we pass on the street.

5.2 Acceptable vs Unacceptable Failure Modes

We consider two possible failure types: misclassification and failure to learn to recognize a person despite frequent encounters. While misclassification is unavoidable, it is less harmful to misclassify a subject as an unknown (false negative) than as another known person (false positive). However, if this continues after subsequent encounters, it translates into the second failure type, which we consider unacceptable.

5.3 Real Time Performance

A design constraint commonly encountered is that the robot must perform at natural interaction rates. Delayed performance may generate confusions as

well as recognition errors. This poses several constraints in our implementation, such as lower image resolution level.

5.4 Prior Assumptions

Studies have found that newborns display a clear visual preference for face like stimuli, suggesting the presence of innate perceptual organization [18]. Taking this finding as a working assumption, we incorporated a face detector ¹ into Kismet's attention system (see implementation section for details).

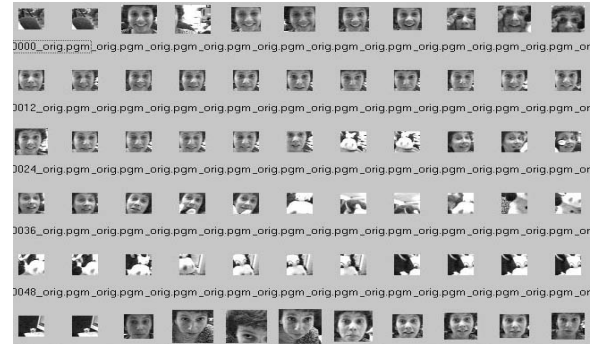


Figure 2. A sample of the face detector [14] output taken from an interaction sequence. Both the camera and target move around, creating a complex and noisy environment. Note the large variations in distance, viewing angle, and lighting.

5.5 Complex Environment

As we know, embodied social interaction translates into a complex, noisy, and constantly changing environment. The system has to work directly from video stream input, instead of nicely cropped images. Both the robot's cameras and the interaction subjects are moving, leading to variations in viewing angle, lighting, distance, etc. Figure 2 illustrates a sample set of detected faces taken from an interaction sequence (cropped based on the outcome of a face detector [14]). This makes both the training and recognition tasks more difficult. In addition, the task complexity increases even more when multiple people interact with the robot concurrently.

5.6 Eigenface Method

As mentioned above, we implemented the face recognition using the eigenface technique [5]. The eigenface algorithm is widely implemented and well known for its simplicity and computational efficiency, an absolute necessity for our real time constraints. The eigenface method uses an information theory approach of coding facial images,

¹ This software was developed by Paul Viola and Michael Jones [16].

where it attempts to find the principal component of the distribution of faces, or the eigenvectors of the covariance matrix of the set of face images. The algorithm can be described as the following steps: 1. Collect a set of characteristic face images of the known individuals, with some variations in expression and lighting condition. 2. Calculate the eigenfaces of the data set (face space) based on each image's top eigenvectors with highest eigenvalues. 3. Each class is represented by averaging its coefficient vectors, which are the projection of its images to the face space. 4. choose thresholds that define the maximum allowable distance from the face class and face space for recognizing new images.

Observation of the recognition system's performance on a random sample of images indicates similar findings to [12] and [13]. Performance is highly dependent on correlation between face alignment in the training and test set. This means that a person's face is only recognized if his/her faces in the training set are in similar orientation. Also, performance degrades in the presence of facial expressions and changes in scale. In [13], subjects were requested to minimize head motion in order to ensure proper alignment. This is not an option in our case because imposing restrictions on subjects' posture and expression will greatly restrict social interactions with the robot.

5.7 Online, Unsupervised, and Empty Training Set

[5] suggested that the concept of face space in the eigenface method allows the ability to learn and subsequently recognize new faces in an unsupervised manner, starting from an initial training set. As mentioned before, we are particularly concerned with the shortcoming of storing a set of manually labeled faces for the robot to learn without grounding them to direct sensory experience. This poses a requirement for our recognition system to start with an empty training set, which raises the following issues: 1. Given an empty training set, how do we begin without having any notion of face space? 2. How do we choose thresholds for maximum allowable distance from the face space and known classes without labeled data?

6. Implementation

6.1 System Overview

As shown in figure 3, visual input from the camera is passed as input to the face detector system [14]. Face detector sends information regarding location of faces found within the robot's visual range, if any, for further pre-processing, where the facial image is

cropped, scaled, normalized, and masked. Based on the training set, the pre-processed image is then recognized to generate an initial hypothesis². Each hypothesis is stored in the hypothesis history buffer to determine whether or not the input face is of a known or unknown individual. After the recognition process, post-processing may be performed on the test image in order to improve future recognition.

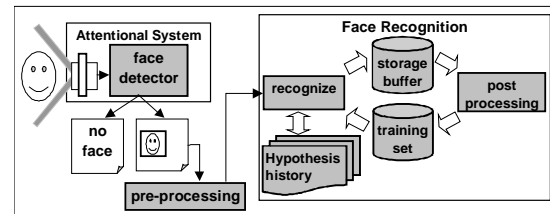


Figure 3. The schematic of the online and unsupervised face recognition system. The system receives video stream as input and learn to recognize individuals in an online and unsupervised manner.

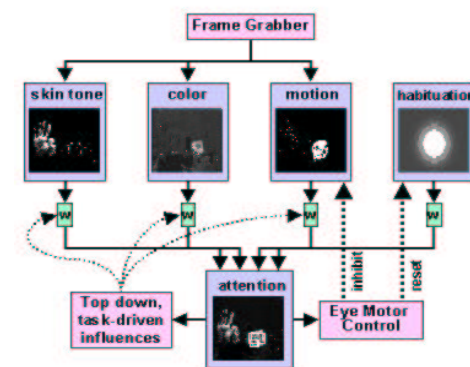


Figure 4. Kismet's visual attention system [15] picks out low-level perceptual stimuli (highly saturated colors, motion, face, and skin tone) that are particularly salient and direct the robot's attention to gaze toward them.

6.2 Attentional System: An Interface to the Environment

Kismet's attention system acts to direct computational and behavioral resources toward salient stimuli and to organize subsequent behavior around them [15]. As shown in figure 4, the pre-attentive system processes information about basic visual features across the entire visual field. For Kismet, these bottom-up features include highly saturated color, motion, face, and colors representative of skin tone. The habituation influence provides Kismet with a

² The face recognition was adapted from Turk and Pentland's eigenface-based face recognition system [5].

primitive attention span. All four factors influence the direction of Kismet's gaze. This greatly simplifies the interface between the face recognition system and the environment. The face recognition system is simply activated when a face is detected within the visual field. If the face is the most salient stimulus at the time, the cameras will track it, maintaining it within the visual field.

6.2.1 Face Detector

The task of the face detector is to determine whether a face exists within the robot's visual field at all times. Given periodic camera input, the face detector outputs the location of existing faces, if any. Clearly, a robust, accurate, and fast face detector is key in this implementation. We used a frontal face detection system developed by [14] which satisfies both the performance and real-time criteria on our slightly different environment: 128 by 128 pixel images on 800 MHz PCs.



Figure 5. A set of input images after preprocessing (scaling, normalizing, aligning, and masking).

6.3 Pre-processing

In order to minimize variations generated by the noise in the environment, each face image found within the visual range is pre-processed. Since faces are detected at different distances, each image is scaled to a 40x40 greyscale pixel image. A simple alignment procedure is applied to roughly correct off-center faces (only effective on frontal images). Each pixel in the image is normalized in order to alleviate lighting variations. Lastly, each face image is masked using 2-D gaussian, thereby diminishing background variations (See figure 5).

6.4 Face Recognition

Given an input stream of potential face images of people interacting with the robot, the system has to perform two tasks: discard non-face images and determine which individual the input face belongs to, if known. Since no training set is available initially (pre-training phase), the system essentially collects and analyzes input data in order to extract a set of

characteristic images per individual in the input set and generate a training set. Once a training set is formed, the online training phase begins.

6.4.1 Pre-training Phase

In order to perform its task, the recognition system must first acquire some idea about the distribution of face space and individual face classes. Thus, the system silently observes while the robot interacts with people. Much like the cross validation method, the system collects a batch of input data, iteratively holds out each image and performs step 2 and 3 of the eigenface algorithm (treating each image as a class) on the remaining images in the batch. This procedure is repeated for each batch of input.

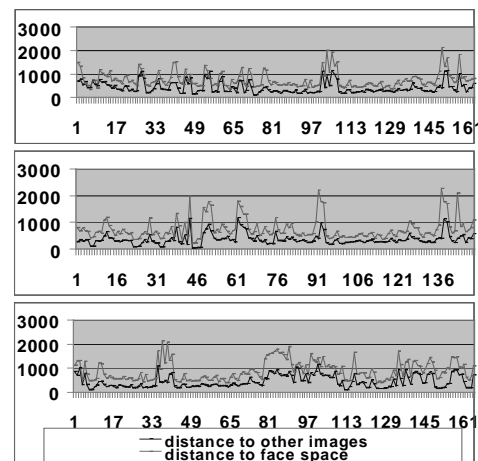


Figure 6. Data analysis during pre-training phase. The x-axis represents each image in an input batch, taken from an interaction sequence. The y-axis represents each image's distance to the estimated face space and distance to its closest neighbor.

Figure 6 shows each input's distance to roughly estimated face space and distance to its closest neighbor, calculated from several initial batches. Each batch contains roughly 200 images. The system observes these results to decide on threshold values for the maximum allowable distance to the face space (Ds) and a known face class (Dc). Assuming that most of the input images contain faces, Ds was chosen to be 1800 to include most of the data points. Histogram analysis of each input's distance to its closest neighbor; Dc was chosen to be 400. Using these threshold values, non-face images are immediately discarded. The system then determines for each input image, all other images that are within the allowable distance to it. Clustering is performed using a simple algorithm: if A is close enough to B and A belongs to cluster x, then B is placed into cluster x. Once a cluster reaches a large enough size, it is placed in the training set as a new class.

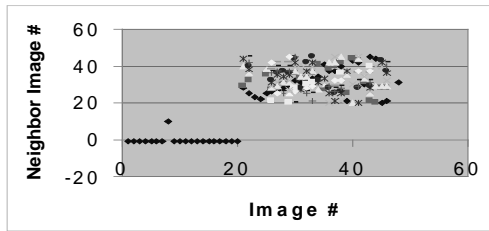


Figure 7. A sample clustering process with 2 individuals in an input set. Image 1-20 belongs to person 1. Image 21-50 belongs to person 2. The x-axis represents each image and the y-axis represents other images in the set that are close enough to it ($y = -1$ if none are close enough).

Figure 7 illustrates a sample clustering process, starting with two individuals' faces in a batch of images. The clustering process places 25 images of person 2 into a cluster and 2 images of person 1 into another cluster. This means that person 1 is poorly represented and thus, more images will be needed for the system to be able to train on person 1's images. Note that the clustering process works without knowing how many individuals there are in an input batch. Thus, we do not need to notify the recognizer whenever an individual comes and leaves, allowing the recognizer to perform in an automatic and unsupervised manner.

6.4.2 Online Training Phase

Initially, the training set is quite small, consisting of only a few individuals. At this point, the system incrementally improves its knowledge of the face space. Each input image is projected into the eigenface basis and the distance between its coefficient vector to the face space and each known class is calculated. Image is immediately discarded if its distance to the face space $> D_s$. If its distance to a known class $< D_c$, it is classified as the corresponding individual. Otherwise, it is perceived as an unknown. Similar to the pre-training phase, each unknown image is collected in batch for post-processing, where the same clustering procedure is applied (distributed across several processors).

Keep in mind that the recognizer is still in the learning process and frequent misclassifications are expected. Thus, these batches may contain images of new individuals, known individuals, or both. If a large enough cluster is found, it goes through another clustering process where each image in the cluster is tested against the existing training set. The cluster is then either added as a new individual or into an existing class depending on how far it is from known individuals.

This method allows the system to learn to recognize new individuals and incrementally improve its training images over time. The more variations in expressions, lighting, and pose there are in an individual's training set, the better the recognition performance is. Thus, the more one interacts with the robot, the more likely it is for the system to obtain their snapshots with various expressions, pose, etc. However, if one just happens to pass by, they would be represented poorly in the database and thus, not easily recognized.

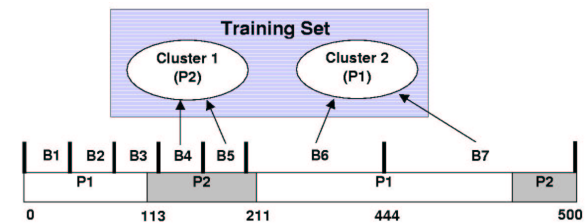


Figure 8. The beginning of the online training phase. P1 and P2 interact with the robot during discontinuous sessions. Initially both P1 and P2 are unknown. Batches (B_i) of 40 images are collected for post-processing. If a large enough cluster is found within a batch, it is placed into the training set.

The online training phase is done with the same 2 individuals in the previous phase, interacting with the robot in discontinuous sessions (see figure 8). Initially, both individuals are unknown. No clusters are made from the first three batches. The first cluster consisting of P2 is made from batch 4. Batch 5 adds to this cluster. Person 1 then comes back and the next two batches (B_6 and B_7) are used to create a cluster for person 1. Note that it takes longer for the system to form a cluster for person 1. Later observations indicate that the nature of one's interaction largely affects how the system behaves. We discuss this issue in more details in the following sections.

7. Experimental Result

The goal of this experiment is to evaluate the performance of the online face recognition system using data acquired from natural interaction with both caretakers and naive subjects. We would like to examine the range of behavior exhibited by people during interactions. We also hope to investigate system behavior, addressing issues such as how much the system performance varies among individuals. Seven subjects (including the two individuals in pre-training phase) were asked to play with the robot in several discontinuous sessions. Toys were provided and interaction time was not regulated. Each subject ended up spending different amount of time with

Kismet. All interactions are one on one. However, pause in between sessions is not necessary, meaning that the system should be able to handle concurrent interactions with the constraint that only one person is salient at a time.

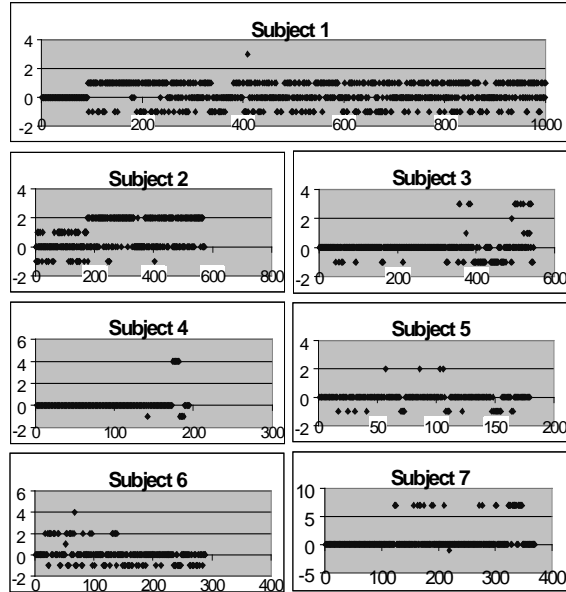


Figure 9. Experimental Results. The x-axis represents time. The y-axis represents the system's output, $y = -1$ (non-face), 0 (unknown), or a known class.

Figure 9 illustrates the output of the recognition system for each person, concatenated over several experiment sessions. The system was able to learn about subject 2, 3, 4, and 7. Subject 5 and 6 were never learned, meaning their images were never put in the training set. Both only spent a short amount of time with Kismet. A few number of false negative errors occurred during subject 1 and 3's session. No false positive learning took place, meaning the system never clusters one individual into multiple clusters and vice versa. As shown in table 1, a few non-faces and badly cropped images were mistakenly placed into the training set, but no errors were made, i.e. no individuals were placed in another individual's training set.

# images	S1	S2	S3	S4	S5	S6	S7
Total	1168	572	546	195	179	289	370
Training set	64	96	16	25	0	0	47
Junk in training set	1	2	5	0	n/a	n/a	1
Error in training set	0	0	0	0	n/a	n/a	0

Table 1. Number of images received and resulting train set size per individual. Junk images include non-faces and badly cropped images. Error in training set means inclusion of another individual in one's train set.

8. Discussion

It is important that no false positive learning took place, because assigning multiple individuals into a cluster or vice versa would lead to future confusion when the robot tries to learn other relevant information about those individuals. Experimental results show that the system exhibits a potential to incrementally learn to recognize a few people's faces. More extensive testing is to be done on more subjects to further verify this claim.

We notice that the system behaves very differently across individuals. It takes a long time to collect training images for certain people. Observations of interaction sessions indicate that there is a large variation in the way people interacts with the robot. Some people like to move around to see if Kismet will be able to track them. Others like to use toys to attract Kismet's attention. These variations greatly impact the behavior of the system. Some individuals may be harder to learn than others because many of their face images may be partially covered by toys.

There are a few remaining issues to be discussed. Firstly, the system currently does not make any reliable final hypothesis about who is currently in front of the robot. Note that individuals who are not in the training set are often misclassified as a known individual until their clusters are formed. Presumably, once the system's training set is good enough, recognition rate of known individuals should be significantly higher than these misclassification incidents. Secondly, as the system continues to deal with more people, system parameters may have to be adjusted. For example, batch size and threshold values may have to adapt to the size of training set.

9. Future Work

We intend to implement a better face alignment procedure, which should significantly improve recognition performance. Our immediate plan is to implement the ability to correlate simple contextual information about individuals along with their faces. As mentioned above, the robot may learn to correlate its emotional state with the presence of certain individuals. The robot may also learn to remember a favorite toy that an individual often uses or words that an individual often says while playing with it. Moreover, we plan to extend this work to include other modalities of person recognition. We anticipate that speaker recognition, sound localization, and gaze detection will increase person identification performance.

10. Conclusion

The ability to distinguish among different individuals is crucial for all social animals. [3] identified the following as the basis of complex social interaction: the ability to distinguish conspecific group members as individuals and as kin, remember their relative ranks and past affiliations, and in some case, remember even the personal histories of help given and received from others. We have implemented an online and unsupervised face recognition system to address the issue of how one may implement such social competence in a humanoid robot. Experiments have been performed and results indicate the system is capable of learning to recognize a few individuals interacting with the robot. Lessons about the resulting interactions are discussed, with respect to developing personal relationships between the robot and its caregivers.

11. Acknowledgements

This work was funded by DARPA as part of the "Natural Tasking of Robots Based on Human Interaction Cues" project under contract number DABT 63-00-C-10102. The author gratefully acknowledges Paul Viola and Michael Jones for their assistance in porting their face detector to Kismet. We would also like to acknowledge usage and include the copyright notice of Turk and Pentland's face recognition system (Copyright 1992, Massachusetts Institute of Technology. All Rights Reserved).

References

- [1] C. Breazeal, Sociable Machines: Expressive Social Exchange Between Humans and Robots. Sc.D. dissertation, Department of Electrical Engineering and Computer Science, MIT (2000).
- [2] P. Carroll, Sociability and Intelligence. [Http: //nua-tech.com/paddy/ethology.shtml](http://nua-tech.com/paddy/ethology.shtml) (1999).
- [3] R.W. Byrne, Human Cognitive Evolution, Chapter 4 in *The Descent of Mind*, ed. M. Corballis and S. Lea. Oxford University Press (2000).
- [4] W. Zhao, R. Chellappa, A. Rosenfeld, P. Phillips, *Face Recognition: A Literature Survey*.
- [5] M. Turk, A. Pentland, Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, Vol. 3 No. 1, pp. 71-86 (1991).
- [6] K. Dautenhahn, Getting to Know Each Other - Artificial Social Intelligence for Autonomous Robots. *Robotics and Autonomous Systems* 16, pp. 333-356 (1995).
- [7] R. Brunelli, D. Falavigna, Person Identification Using Multiple Cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-17, No. 10 (1995).
- [8] J. Kittler, Y. Li, J. Matas, M. Ramos Sanchez, Combining Evidence in Multimodal Personal Identity Recognition Systems. *International Conference on Audio and Video-based Biometric Person Authentication*, Switzerland (1997).
- [9] C. Nakajima, M. Pontil, B. Heisele, T. Poggio, People Recognition in Image Sequences by Supervised Learning. A.I. Memo No. 1688, C.B.C.L. Paper No. 188. MIT (2000).
- [10] J. Weng, C. Evans, W. Hwang, An Incremental Learning Method for Face Recognition under Continuous Video Stream. *Fourth International Conference on Automatic Face and Gesture Recognition*, Grenoble, France (2000).
- [11] A. Pentland, T. Choudhury, Personalizing Smart Environments: Face Recognition for Human Interaction. *IEEE Computer*. Special issue on Biometrics (2000).
- [12] G. Sukthankar, Face Recognition: A Critical Look at Biologically-Inspired Approaches. [Http: //www.cs.cmu.edu/~gitar/16-721/final/final.html](http://www.cs.cmu.edu/~gitar/16-721/final/final.html) (1999).
- [13] Y. Yacoob, H. Lam, L. Davis, Recognizing Faces With Expression. *International Workshop on Automatic Face and Gesture Recognition*, Zurich (1995).
- [14] P. Viola, M. Jones, Robust Real-time Object Detection. Technical Report Series, CRL 2001/01. Cambridge Research Laboratory (2001).
- [15] C. Breazeal, B. Scasselati, A Context-dependent Attention System for a Social Robot. *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, pp. 1146-1151. Stockholm, Sweden (1999).
- [16] S. Eickeler, F. Wallhoff, U. Iurgel, G. Rigoll, Content-based Indexing of Images and Videos using Face Detection and Recognition Methods. *IEEE Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Salt Lake City, Utah (2001).
- [17] T. Field, D. Cohen, R. Garcia, R. Greenberg, Mother-Stranger Face Discrimination by the Newborn. *Infant Behavior and Development* 7: 19-25 (1984).
- [18] E. Valenza, F. Simion, V.M. Cassia. Face Preference at Birth. *Journal of Experimental Psychology, Human Perception and Performance* 22:892-903 (1996).
- [19] S. Furui, An Overview of Speaker Recognition Technology. *ESCA Workshop on Automatic Speaker Recognition Identification Verification*, Switzerland, 1-9 (1994).